DIA2: Web-based Cyberinfrastructure for Visual Analysis of Funding Portfolios

Krishna Madhavan, *Member, IEEE*, Niklas Elmqvist, *Senior Member, IEEE*, Mihaela Vorvoreanu, Xin Chen, *Student Member, IEEE*, Yuetling Wong, Hanjun Xian, Zhihua Dong, Aditya Johri

Abstract—We present a design study of the Deep Insights Anywhere, Anytime (DIA2) platform, a web-based visual analytics system that allows program managers and academic staff at the U.S. National Science Foundation to search, view, and analyze their research funding portfolio. The goal of this system is to facilitate users' understanding of both past and currently active research awards in order to make more informed decisions of their future funding. This user group is characterized by high domain expertise yet not necessarily high literacy in visualization and visual analytics—they are essentially casual experts—and thus require careful visual and information design, including adhering to user experience standards, providing a self-instructive interface, and progressively refining visualizations to minimize complexity. We discuss the challenges of designing a system for casual experts and highlight how we addressed this issue by modeling the organizational structure and workflows of the NSF within our system. We discuss each stage of the design process, starting with formative interviews, prototypes, and finally live deployments and evaluation with stakeholders.

Index Terms—visual analytics, portfolio mining, web-based visualization, casual visualization, design study

1 Introduction

As visual analytics technologies gain widespread adoption across a broad array of disciplines, the visual analytics community increasingly finds itself catering to an entirely new brand of users. One such population is highly qualified professionals that are experts in their fields, yet possess little knowledge of visualization and visual analytics. Their dynamic work environment also leaves them with little time or opportunity to learn new systems. Unlike the previously proposed definition of casual visualization, which provides visualization to casual users driven by personal goals and motivations [32], we call this new brand of users *casual experts* given their extensive expertise in a domain but a casual approach to visual analytics methods. We believe that the visualization needs of such users can be best met by a design study approach [35] that investigates and understands their approach to problem solving in their domain of expertise and adapting that to the design of the visualization. We want to emphasize though that the word "casual" in casual expert refers only to users' attitude towards visualizations and not their domain work which is often very high-stakes.

In this paper, we present a design study of a *web-based visual analytics* platform called **DIA2** (Deep Insights Anytime, Anywhere) designed for this new brand of casual experts. The DIA2 system is a knowledge mining platform for portfolio management [22] of awards made by the U.S. National Science Foundation (NSF). It enables program managers and professional staff at the NSF to view and analyze the projects, publications, and people involved in past and currently active NSF awards. The intended audience of DIA2 perfectly embodies the casual experts moniker previously discussed: DIA2 users are academics with a high degree of training in their discipline, yet with little to no training and interest in advanced visualization and analytics. In keeping with the spirit of such casual experts, the design philosophy of the DIA2 project is "no manuals, no training." Instead, any training necessary in using DIA2 is designed to happen during the user experience of performing the intended tasks through several mechanisms: (1) strict adherence to norms and standards used in graphical user interface design, (2) clear visual affordances and labeling, and (3) *progressive refinement* of all visual representations where visual complexity is only gradually added in response to direct and reversible actions performed by the user. The intention with the progressive refinement mechanism is that every new visual state for a particular visualization, including the first one, should be easily comprehensible to a casual expert. To convey complex data, the user would iteratively and interactively query the visualization to gradually add this complexity.

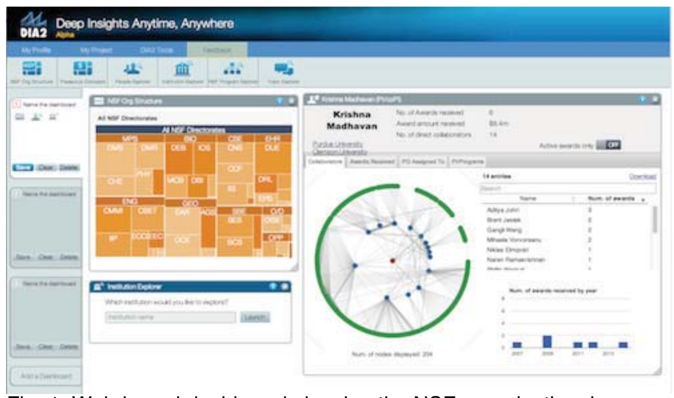

Fig. 1. Web-based dashboard showing the NSF organizational structure using a treemap (left) as well as a Person Profile (right) in the DIA2 platform.

The DIA2 system (the public version of the website is accessible at http://www.dia2.org/) is a web-based interface for a large-scale online database and uses a visual dashboard design (Fig. 1). Users create data widgets on the dashboard canvas, and widgets can then be freely moved, resized, and deleted. Dashboards are persistent across sessions, and users can create and name several dashboards for different purposes. Each data widget is interactive and combines visual representations and underlying data tables for different purposes. DIA2 currently supports widgets for exploring the NSF organizational structure, concepts and keywords, investigators, institutions, research programs, and research topics. The visual representations used include treemaps, mosaic plots, ego-networks, and various statistical graphics such as bar charts, pie charts, and time-series plots, all of them implemented using the progressive refinement design guideline discussed above. Furthermore, the system supports advanced search features tying the widgets together.

The primary contribution of this design study is the goal-directed design process [9] we followed in creating the DIA2 platform. The design process started with gathering in-depth information in order to create personas of users inside the National Science Foundation.

- *Krishna Madhavan, Niklas Elmqvist, Mihaela Vorvoreanu, Xin Chen, and Yuetling Wong are with Purdue University. E-mail: {cm, elm, mihaela, chen654, wong64}@purdue.edu.*
- *Hanjun Xian and Zhihua Dong are with Microsoft Corporation. E-mail: {hxian, zhdong}@microsoft.com.*
- *Aditya Johri is with George Mason University. E-mail: ajohri3@gmu.edu.*

- *Submitted to IEEE VAST 2014. Please do not redistribute.*

Manuscript received 31 Mar. 2014; accepted 1 Aug. 2014. Date of publication 11 Aug. 2014; date of current version 9 Nov. 2014. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org. Digital Object Identifier 10.1109/TVCG.2014.2346747

1045-9219 © 2014 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

We attempted to understand not only their daily tasks and needs, but also higher level goals, as recommended in [9]. We conducted a comprehensive set of focus group sessions and individual interviews with program officers (POs) and science assistants (SAs) at the NSF. These interviews led us to derive the concept of casual experts. Based on personas and design requirements we used sketches and wireframes to conceptualize and evaluate initial designs and then deployed a live alpha version of the DIA2 system internally at the NSF. This version was evaluated with members of our user groups. In this paper, we report on every stage of the research project and review results from our evaluation studies. We close the paper with implications for designing for casual experts and plans for future work.

2 Background

Our work in this paper combines ideas from scientometric visualization, visual analytics for organizations, and new ideas on design study methodology. We review these research topics below.

2.1 Visualizing Research

Scientometrics is the study of measuring, analyzing and discovering science growth, structure, interrelationships and productivity [20]. It has overlapping interests with bibliometrics and informetrics. As a result, scientometrics research is often done using bibliographic visualization tools. These tools include BIVTECI [24], a prototype system proposing a minimum set of functions necessary for effective bibliography visualization, Butterfly [21], a system providing a 3D information visualizer for assessing DIALOG's Science Citation database using a so-called "organic user interface," and CiteSpace II [8], which visualizes co-authorship and co-citation relationships.

DIA2 is an analytics platform for searching, viewing, and analyzing the NSF research portfolio for casual experts. It has many features in common with the scientometrics and bibliometrics research such as most of the data is related to scientific awards, research and publications, the personal collaboration network is similar to the co-authorships in bibliometrics, and they both represent and predict cutting-edge research trends. Therefore, such bibliographic visualization techniques can also be utilized in DIA2.

A particularly relevant effort is the MultiNode-Explorer [13], a visual analytics framework that generates web-based multimodal graph visualization from multidimensional data. It accepts an entity-relationship schema transformed from the multidimensional data, a set of relational data tables, an interface specification file, and visualizes the data as node-link diagrams. As the NSF portfolio datasets are multidimensional and multivariate, the MultiNode-Explorer framework is a useful reference implementation for our visualization process, with the important caveat that DIA2 needs additional visual representations for its multifaceted and multidimensional datasets rather than just node-link diagrams.

2.2 Visual Analytics for Organizations

Several papers in HCI have documented the obstacles encountered by large companies when conducting interface design, evaluation, and usability testing (e.g., inability of interface designers to obtain access to users, resistance to iterative design, and lack of communication) [1, 18, 31]. This previous work mainly focuses on designing products for customers rather than building visual analytics tools for internal employees. Sedlmair et al. [34] extend this work by documenting the challenges encountered by visualization researchers when designing for internal employees of large companies. They point out that the workflow, bureaucracy, and hierarchical structures could all pose challenges to the design and evaluation process. All the above-mentioned studies happen in industry settings. In this paper, we are designing for a federal government research funding agency: the U.S. National Science Foundation. NSF has very different work practices and culture compared with industrial companies, and the problems that workers need to address in their everyday work are unique to this context. Yet, common across settings is the restricted mental capacity of users to be able to pay attention to information, including visualizations.

There is general agreement among visual analytics scholars that humans are parsimonious problem solvers [15, 16]. Consequently, they frequently choose the simplest heuristics that are available to them and are adequate for a given task. Therefore, a critical task of visual analytics designers is to present information to users within a relevant context to mitigate the problem of cognitive load. In particular, for visualizations that are complex and contain numerous semantic data points, being able to leverage existing heuristics or mental models is a distinct advantage. In spite of the usefulness of considering human factors in visual analytics design, Tory and Möller [37] suggest that human factors are often neglected in visualization systems. They argue that "more attention should be paid to users who must view and manipulate the data because how humans perceive, think about, and interact with images will affect their understanding of information presented visually" [37] (pg. 72). Furthermore, they argue that there is a lack of focus on visually displaying users' mental models and on helping users improve their mental models. The authors suggest that visual systems can help users by providing ways to organize and share ideas. Liu & Stasko [28] look specifically at the role of mental models in visualizations and argue that although there has been some emphasis within the field on internal cognitive mechanisms, there is a need to account for ecological and situated accounts of cognitive behavior. They review the broad literature on mental models and argue that a mental model allows preservation of schematic, semantic or item specific information and is beneficial as it allows construction and simulation of a problem in working memory, thereby aiding reasoning. The importance of considering human factors and their mental models provides the conceptual foundations for our design philosophy. In the formative research phase, we aimed to capture users' mental models related to how they work with and report data.

3 Design Study Approach

Sedlmair et al. define a design study as "a project in which visualization researchers analyze a specific real-world problem faced by domain experts, design a visualization system that supports solving this problem, validate the design, and reflect about lessons learned in order to refine visualization design guidelines" [35] (pg. 2431). Design studies do not seek to create new visualizations; rather, they seek to solve real-word problems and provide transferable guidelines on solving such problems through reflection. Compared to technique-driven visualization research, design studies are one approach of problem-driven research. Although many design study papers have appeared in recent years (e.g. [14, 30, 33, 40]), studies that design visual analytic systems for organizations such as a government funding agency are still rare and are therefore highly valuable to the design study knowledge pool.

4 Context: U.S. National Science Foundation

According to its website, The U.S. National Science Foundation (NSF) is an independent federal agency with a total workforce of about 2,100 at its Arlington, VA, headquarters. This includes approximately 1,400 career employees, 200 scientists from research institutions on temporary duty, 450 contract workers, and the staff of the National Science Board (NSB) office and the Office of the Inspector General. The NSF leadership has two major components: a director who oversees NSF staff and management responsible for program creation and administration, merit review, planning, budget and day-to-day operations; and a 24-member NSB of eminent individuals that meets six times a year to establish the overall policies of the foundation. The director and all Board members serve six-year terms. Each of them, as well as the NSF deputy director, is appointed by the President of the United States and confirmed by the U.S. Senate.

NSF was created by Congress in 1950 "to promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense…" With an annual budget of about $7.0 billion (FY 2012), NSF supports approximately 20 percent of all federally supported basic research conducted by America's colleges and universities. In many fields such as mathematics, computer science and the social sciences, NSF is the major source of federal funding for researchers and educators. NSF works to ensure that research is fully integrated with education to support the training of tomorrow's scientific and engineering workforce. NSF keeps track of current research and maintains constant contact with the research community to keep abreast of the latest ideas, and to choose the most promising people to conduct the research.

Each year NSF receives approximately 40,000 proposals of which approximately 11,000 are funded. Program officers working at NSF are responsible for the selection of proposals with the highest merit and they utilize review panels to evaluate submitted proposals. In order to be able to put together the panel with the right expertise, they need information about other experts in the field; they need to figure out conflicts of interest among proposal authors and panelists, if any; and, they need to understand the importance of an idea for the field beyond the review provided by experts, in particular to avoid duplicate funding. All these tasks require significant knowledge as well as the ability to quickly derive new insights from existing data. This is the primary need we address with our system.

5 METHODS

As we approached this project, our focus was on gaining a solid understanding of users' goals, needs and workflows, which would help us identify their mental models of working with data and reports. We went into the design project with a "blank slate" attitude ready to learn as much as we could about our users before creating any solutions.

To accomplish this goal, we followed Cooper's [9] goal-directed design methodology. We gained access inside the NSF and conducted nine focus groups over two separate visits with 31 NSF personnel that resulted in about eight hours of audio recording. The access constraints placed by the organization were a research challenge we had to cope with. For example, it was not possible to conduct observations or contextual inquiry [19], and we had little to no control over sampling. To compensate for these limitations, we relied on detailed and intensive interviewing. We analyzed the qualitative data using the method recommended by Cooper [9] that seeks to identify similar behavior patterns that form the basis for creating personas. The method is somewhat similar to thematic analysis [4, 5] and the open and axial coding steps of grounded theory [7], the only difference being that the focus is on identifying and grouping patterns of behavior and higher order goals. Three personas, described in the next section, emerged from the initial user research.

5.1 Personas

Research with users inside the NSF revealed categories of users whose existence we were not even aware of. As we went into the research, we assumed program officers would be the main user group. However, three different user groups emerged from the ethnographic interviews we conducted inside the NSF. We created one persona for each user group. Cooper [9] defines personas as "composite archetypes based on behavioral data gathered from actual users" (p. 76). Personas are useful design tools because they can help designers "develop an understanding of our users' goals in specific contexts" (p. 76) as opposed to an abstract understanding facilitated by impersonal demographic information. Personas usually have a name, a photo, an explanation of the person's goals, work context, as well as needs and frustrations related to the aspect of work we design for. Three personas emerged from our formative user research:

James - Program Officer (PO). James' main responsibilities are to oversee and manage research funding. He is involved in authoring calls for proposals, organizes review panels that evaluate submitted proposals, and oversees funded projects. He is often asked to prepare reports about the state of funding and relies on science assistants to find and analyze the needed data. These reports have to account for the specific organizational division James works in and its budget. However, James is interested in a broader view of work funded by other divisions. Informal conversations with colleagues have revealed that similar topics are funded by different divisions, but there is no systematic way to document and access organizational history. James holds a PhD and joined the NSF because he is committed to advancing research in his discipline. However, he finds that most of his work day is spent in "fire-fighting" tasks that leave insufficient time for reflection on the broad research directions of his discipline.

Amy - Science Assistant (SA). Amy has recently graduated with her MS degree. She is employed by NSF for a period limited to two years to help POs directly with data retrieval and analysis related to numerous aspects of their work. She is responsible for helping a number of POs in specific organizational divisions. Even though Amy is highly qualified, she spends most of her day acting like a human search engine, manually parsing search results from databases that are difficult to query. It might take Amy as long as two weeks to create a report to her assigned program officer, and she cannot do so without a lot of manual work and help from other science assistants and program officers who need to review and validate her query results before she can compile any data from them. Amy is on the front lines of working with NSF's data and as such is intimately familiar with the NSF's complex ways of classifying information by proposal type and organizational division. She has experimented with a few other new tools for interfacing with the data, but does not trust them easily. She usually double-checks against older, more trusted tools to ensure the results are accurate. In many ways, Amy is the persona who will review new systems most critically and with the highest attention to detail. The team did not know about the existence of science assistants before conducting initial user research.

Matt – Rotator. Matt is a recently tenured associate professor who is serving as a temporary program officer at NSF for a period of two years. NSF employs rotators on a regular basis as part of the organization's philosophy to bring fresh and relevant perspectives to bear upon its research agenda. Matt's biggest challenge is to gain an understanding of the funding portfolio he has inherited and is now in charge of managing. It takes him months to understand the nature of the awards in his portfolio before he can become fully informed and productive. Matt is the most vulnerable of the three personas, as his needs for information are dire and he does not benefit from historical knowledge and a rich social network as the POs do.

We realized that our design needed to make sense to Matt and help him overcome initial barriers by learning at a glance about the state of his funding portfolio, but needed to pass the careful scrutiny of detail-oriented science assistants.

5.2 Formative Design: Casual Experts

The term that emerged to describe all of our user groups was *casual experts*, a concept we derive from Pousman et al.'s [32] work on *casual visualization*. All three personas had advanced domain-specific expertise, but, with the exception of a few science assistants, little expertise in information retrieval and no expertise in information visualization (for example, many of the program officers do not have a background in a data-centric discipline). This distinguishes our casual experts from traditional expert users, who—through training and prior knowledge—are knowledgeable in data science and methods. Moreover, our personas had little time and inclination to learn new visual analytics and visualization tools. However, unlike true casual users [32], our casual experts do have a professional interest and motivation in using a tool if it can help them perform their duties faster and more efficiently. This pointed to the need to design a tool that is as easy to learn and use as casual visualizations, but offers accurate and trustworthy insights that can be used to support important organizational decisions. If the new system we designed required training, it would not be used; in fact, we learned of some such new systems had been introduced in the past and fallen by the wayside.

The design requirements that emerged from the formative design had therefore to take into consideration the users' needs to access and assess information at a glance, while keeping it within the strict boundaries of NSF's organizational structure which was heavily reflected in users' mental models of how they reported data. Getting an at-a-glance overview of funding portfolios emerged as the main design requirement. Users in all three groups emphasized the need to see, at a glance, how their organizational unit's funds were invested.

A second design requirement was to follow rigidly the NSF's organizational structure. It became apparent that users' mental models reflected the organizational structure. We identified the types of information needs and reports that users needed to generate (e.g. funding rates) and noticed that each one of them was dependent upon a specific organizational unit such as program, division, or code.

Third, it became apparent that internal organizational language was used rigidly and very specifically. We made an effort to learn this language and apply it to labels on the interface we designed. For example, users inside the NSF differentiate between proposals and awards, and define "awardee" as an institution, not an individual. Even though the language, organization, and work culture of the NSF were initially foreign to us, we made an effort to learn them quickly, represent them in the system design and user interface, and improve them based on continuous user feedback.

When looking at the work practices of our users, several other aspects stood out that further influenced the design of DIA2. For example, we were confronted with the array of systems that POs and SAs had to use to be able to get the required information. NSF has over the years acquired numerous information systems that do not necessarily interact or integrate with each other. Therefore, there is no one 'place' to go to in order to find solutions to a problem. Furthermore, many of these systems might have access to the data but more crucially–from the standpoint of our design–these systems were not designed with the users in mind. The systems are largely software packages available commercially or designed in-house by contractors who primarily drew on their experience in the business community. They are not tailored towards the users from a visual analytics and visualization perspective. Even though the systems had the required information or data and could provide them to the users, the presentation was not designed optimally and resulted in little to no use by personnel who were not specifically trained to understand those systems. Furthermore, there was a persistent gap between the expertise of the users–POs and SAs–and the system designers (who were software developers with no understanding of the research context in which POs and SAs operate) that resulted in systems that were hard to use. There is another complication with rotators who are in a temporary position–at loan from their home institutions for a period of 2-3 years–and have to learn the numerous systems in order to be able to complete their work. When they leave, new people have to be trained.

We articulated our design philosophy as "No manuals, no training" to remind the team of the focus on ease of learning. The design goal was for the interface to be easy to learn, and for the visualizations to make sense to our users at a glance.

5.3 Conceptual Design

Because of our focus on ease of learning, we selected the persona with the least amount of organizational knowledge as the primary one to design for. In this case, this was Matt the rotator. We reasoned that any design solution that would satisfy this persona would also serve the other two. At the same time, we kept in mind that science assistants and permanent program officers look at information with a higher level of scrutiny informed by their tenure and experience inside NSF, making them the more demanding user groups.

We then developed use cases for Matt with the help of a team member who had served as a rotator at NSF in the past. The use cases were guided by the question of what a new NSF employee would want to know in order to get up to speed with his or her portfolio of awards. We created lists of types of information the rotator would need to see and then brainstormed solutions for representing this information visually in ways that are easy to learn and understand for our casual expert users.

Members of the user experience team then translated the sketches from the brainstorming sessions into detailed wireframes. The wireframes specified the layout, display, and functioning of each visualization. Special attention was paid to usability guidelines such as Nielsen's [25] 10 heuristics and Norman's concept of affordances [26]. Clearly communicating affordance, or the action enabled by each element in the design, was considered key to creating an interface that would be easy to learn. The various visualization tools were integrated under a dashboard metaphor reminiscent of financial investment dashboards – an idea that emerged from the users' frequent mentioning of the need to get a bird's eye view of their funding portfolios.

Design reviews and cognitive walkthroughs [17] were conducted by the user experience team as sketches and wireframes were developed. A larger cognitive walkthrough session involving all available team members was conducted before launching the alpha version of the tool. We collected informal feedback from users at NSF during the conceptual design phase, but formal evaluation sessions were not possible at that stage, which is a limitation we had to overcome. However, the team was invited to present the system on several occasions in front of various NSF constituencies, who provided feedback upon viewing demos.

An alpha version was made available to users and was evaluated using one-on-one moderated usability interviews. The evaluation results are presented in section 7. The following sections explain the technical design used to support the concepts we developed.

6 DIA2 Technical Design

6.1 Design Rationale

Given that DIA2 users are casual experts, the types of insights they require demand a high level of precision. End users typically use DIA2 to answer critical questions that are part of policy-making decisions affecting a significant number of researchers in the US. Our design choices highlight the need for accuracy, simplicity, and speed that our casual experts repeatedly required. Therefore, our goal is to show how simple visual representations that aid understanding and decision-making are far more critical to casual experts than more complex techniques that may traditionally be considered the state-of-the-art in visual analytics. We make the case in Section 8.2 of this paper that *affordance is innovation*.

Another critical point to note here is that whenever we design an analytics environment for casual experts, there is a risk that we design a "black box" that may make the data opaque to some expert users. Our users wanted a mechanism to show where the data originated from and the ability to crosscheck these data with internal NSF mechanisms. To facilitate this, the DIA2 presentation layer provides users with information about data ranges and the ability to export data as a CSV file into Excel where such comparisons could occur. Only the most proficient users of DIA2 (science assistants) requested this capability. Also, it is important to note that casual experts do not want to truly concern themselves with the algorithms or other technical details. In fact, the reason why DIA2 is being adopted inside the NSF, while many previous efforts have failed, is because we shield our users from the technical aspects of a portfolio mining system while still providing them with mechanisms to crosscheck the results they receive from DIA2.

This design rationale permeates each of the service layers of the system (described in more detail below). DIA2 utilizes an n-tier architecture [23] that treats every layer of the system as a service provisioned to the other layers of the system. Therefore, DIA2 also exhibits all the properties of a service-oriented architecture [3, 12]. DIA2's technical core is designed to provide users with fast response times and high system availability. More specifically, we operationalize these ideas for each of the layers as follows:

- **Data layer**: providing low-level access to all DIA2 data using accurate, high-performance, and effortless mechanisms;
- **Middleware layer**: organizing the DIA2 data in a format that enables aggregated views, search, and visualization; and
- **Presentation layer**: rendering views of the data using familiar visual representations, tables, and lists integrated in multiple user-configurable dashboard layouts.

6.2 Data Layer and Data Mining Approaches

Data Sources: Currently DIA2 archives data from January 1973 to March 2014. Based on user feedback, we elect to expose only the data from 1995 on.. DIA2 uses a combination of structured and unstructured data entities as the base for all the analytic services provided to the end users. The primary database system is a MySQL database that consists of a variety of metadata relevant to NSF grants. In addition, DIA2 also uses full texts of awards abstracts, journal papers, and conference proceedings resulting from a sizeable number of awards, and in many cases actual links harvested from focused crawling [6, 10] of the web for products resulting from NSF awards (such as curricular materials and web resources). DIA2 also includes a warehouse of data derived from surveys conducted as part of NSF program analyses, impact reports generated by the individual NSF programs, and also taxonomies developed by individual programs within the NSF. The total size of the data is 1TB (transactional).

Data Disambiguation: The data layer includes a range of acquisition, aggregation, disambiguation, and completion protocols that ensure data coverage and data cleanliness. The data layer includes a set of algorithms designed specifically for DIA2 implemented via system daemons to continuously evaluate the quality of the data, incrementally request additional data from the various systems inside the NSF, and resolve ambiguity in author names, proposal titles, institutional affiliation, etc.

Data Reduction Approaches: DIA2 utilizes a variety of data reduction approaches to organize its core data. For example, DIA2 uses a combination of *N-gram extraction* and *tf-idf techniques* to extract relevant keyphrases. Keywords denote a single word while keyphrases denote multi-word units. Keyphrases are valuable in describing the content of a single document and provide a kind of semantic metadata and document summary that is useful for a wide variety of purposes. Keywords and keyphrases are particularly useful because they can provide a way of browsing a collection and as a document clustering technique [39]. DIA2 plans to allow folksonomic tagging of documents and data entities, enabling a search taxonomy based on user-supplied keyphrases. Folksonomies developed via user inputs can be extremely valuable in identifying and distinguishing between documents with a high degree of confidence and seeding search effectively.

Ensuring Fast Response Times: DIA2 users repeatedly emphasized speed as a critical requirement in the system. However, when users interact with DIA2, every user click produces a huge demand for data. Traditionally, clicks trigger requests to the database creating a high possibility of a bottleneck at peak data demand. In DIA2 every user click automatically routes the request to the Query Cache Handler (QCH) – which cuts down response times of complex queries by nearly 90%.

Allowing Expert Users Control over Data: The entire data layer is exposed to other parts of DIA2 and to expert users as a set of JSON/RPC services. There are two primary reasons for this architectural decision. (1) Security of the primary data sources is highly critical in systems such as DIA2. The JSON/RPC services allow for better consistency, maintenance, and security of the data components. (2) Secondly, experts in data mining may not want to be constrained by the UI provided to casual expert users and may want to work directly with the raw data for a variety of purposes. This data architecture allows appropriate rationing and control of the data flow out of the DIA2 system while providing standardized data access.

6.3 Middleware Layer

Most of the algorithms, workflow artifacts, and rules that drive various aspects of DIA2 are managed within the middleware layer. DIA2's middleware layer is designed as a set of services that are invoked as needed.

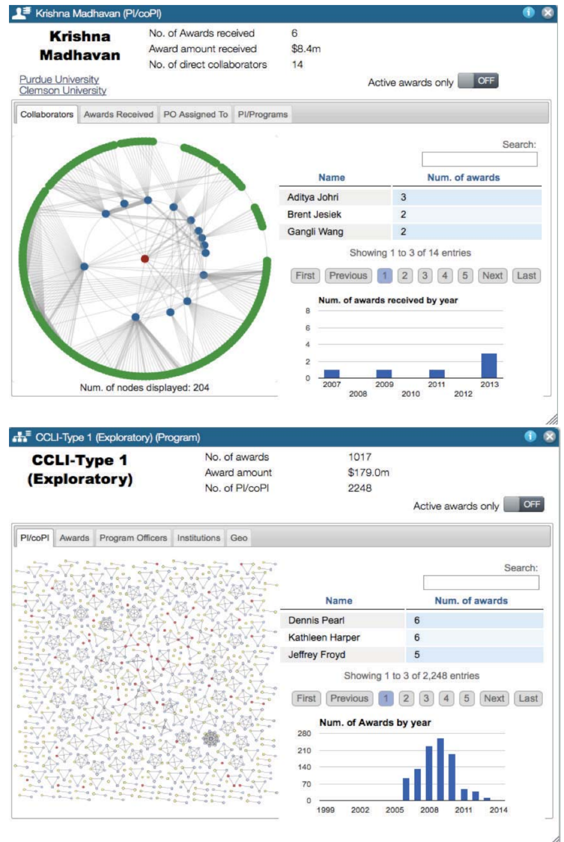

Fig. 2. DIA2 includes a range of services that allow better visual presentation of collaboration data. Nodes represent researchers and links are proposals that have been awarded by NSF. These visualizations also show capacity building within each program (organizational structure).

Rules Services: In DIA2, every data entity is processed based on very specific business rules identified by users. Every view provided to users is driven by these basic business rules that specify, for example, what counts as an award, how to determine the organizational units responsible for award parts, and so on. Furthermore, the rules engine also tracks and provides appropriate processing frameworks to other DIA2 visualization services (such as the treemap visualization) and search aggregation.

While the rules engine determines the appropriate data aggregation and processing framework, visualization services work in coordination with the presentation layer to render the appropriate visual information requested by end users. The methodology used to determine the affordance that a specific visualization provides to the end users is discussed in [29]. As opposed to thinking of individual visualizations as algorithms, DIA2 considers the basic nature of the data to be visualized and creates a service that is generic and abstract enough to serve visualizations specific to data types. We elaborate on this next.

Hierarchical Data Visualization Service (HDVS): DIA2 users are very interested in data that reflect hierarchy. The preferred visual representation of hierarchical data within DIA2 is through the use of treemaps [36]. The HDVS service handles the requests to process organizational structure, programmatic structure, and also taxonomy information within DIA2.

Collaboration Data Visualization Service (CDVS): One of the purposes of DIA2 is to showcase the collaboration networks emerging around individual researchers and organizational entities. To showcase the collaboration around individual researchers, DIA2 utilizes ego-centric social networks, while the organizational structures are visualized using simple flat spring loaded social network layouts. Fig. 2 highlights the type of visualizations provided by the CDVS.

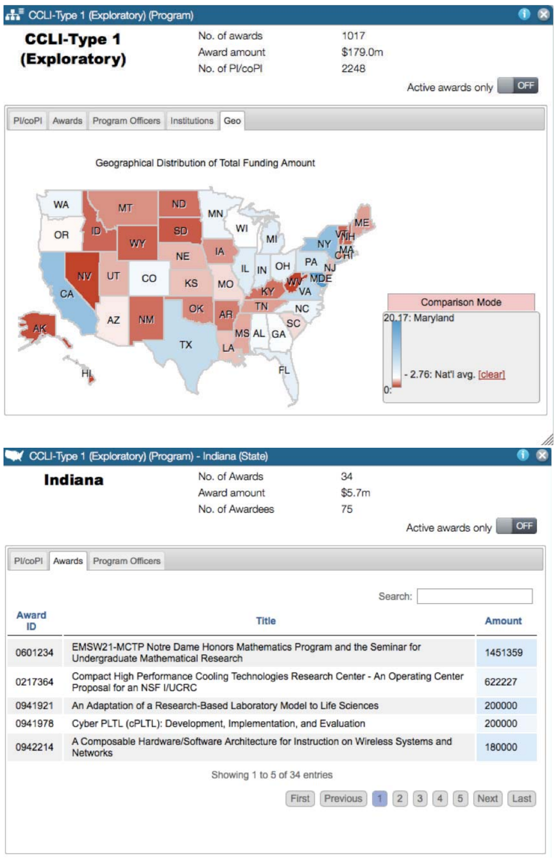

Fig. 3. DIA2 provides a full range of analytics focused on geographical locations. The first graph in the inset is the GDCS working in the “Comparison Mode”.

Geographical Data Consolidation Service (GDCS): DIA2 is at its core a portfolio mining platform. Evaluation of how federal funding is distributed across geographical area emerged from formative research as a critical part of the analytics needed for reports.. To this end, DIA2 is capable of not only providing consolidated data on map overlays, but also drilling down into data aggregation at the level of individual academic institutions within a specific geographical region. Fig. 3 provides an example of the GDCS in action. The GDCS also allows quick comparisons of various data aggregations across the geographical range.

Search Services: DIA2 features the ability to translate any search into a coherent set of analytics. Within DIA2, all data artifacts – people, organizational structures, programs, awards, concepts, keyphrases, and institutions – are searchable. DIA2 utilizes Apache Solr[1] (a derivate of Apache Lucene[2]) to index a wide range of documents. Search services in DIA2 are designed to not return a linear list of results. The aggregated search service continuously interacts with the rules engine and the visualization services to synthesize the results in meaningful ways. Fig. 4 shows a simple search of a concept driving a full set of highly synthesized results. The search results are provided in a simple widget that contains multiple tabs. Each tab has information relevant to one aspect of the search. The search service allows end users to use a variety of Boolean operators to constrain the search results.

6.4 Presentation Layer

The previous sections provided a description of the data and middleware layers respectively. However, these layers are completely hidden from end users. The only aspect of DIA2 that users really interact with is the presentation layer. The presentation layer uses an overall dashboard metaphor, as shown in Fig. 1. The various types of interactive data exploration and visualization tools are individual widgets. The widgets' visual design and interactivity follow the ease of learning imperative and use progressive refinement in order to avoid overwhelming users with too much information.

6.4.1 Widgets and Dashboards

The entire user experience within DIA2 is based on a dashboard metaphor. Users are provided with 3 dashboards (blank canvases) by default with the option to add up to 5 dashboards in any workspace. The limits were determined based on a simulation of resource allocation to enable scaling to a large number of users. All dashboards can be named and saved for future use. In future versions of DIA2, dashboards are also designed to be shareable with other users. Currently DIA2 supports 6 different widgets with 3 more currently being planned. Each widget provides data views with multiple tabs providing different information to the end users. Every aspect of the presentation layer is completely controlled using an XML configurator file. All widgets have a standard descriptor that is packaged along with the code for that particular widget. The descriptor sets and determines the behaviors of the widget and also provides a baseline for the middleware services the widget needs to connect with to provide its functionality.

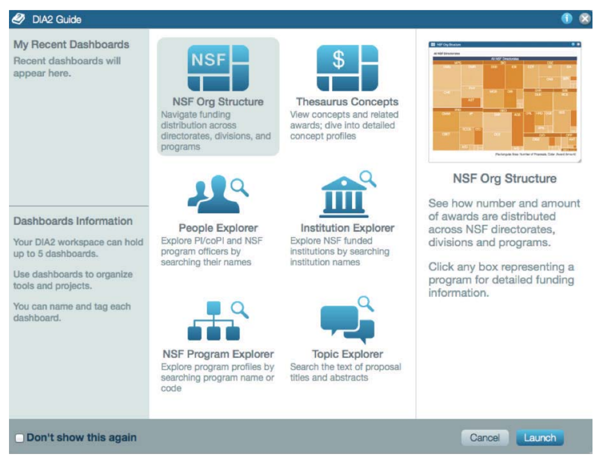

Fig. 5. DIA2 Guide available to users at the launch of the alpha version. The guide provides a simple selector that users can select to learn about each widget and launch it.

As users launch the alpha version of DIA2, they are shown a simple widget selector called the DIA2 Guide (Fig. 5) that provides brief descriptions of each widget. When widgets are opened into the workspace, a small icon showing the status of the widget appears on the dashboard side tab. All dashboards can be saved and cleared. The presentation layer also includes a caching mechanism to speed the rendering of the widgets on users' screens.

6.4.2 Visual Design and Progressive Refinement

Based on our casual expert audience, our design rationale for all visualizations was to choose visual representations that are familiar

[1] http://lucene.apache.org/solr/
[2] http://lucene.apache.org

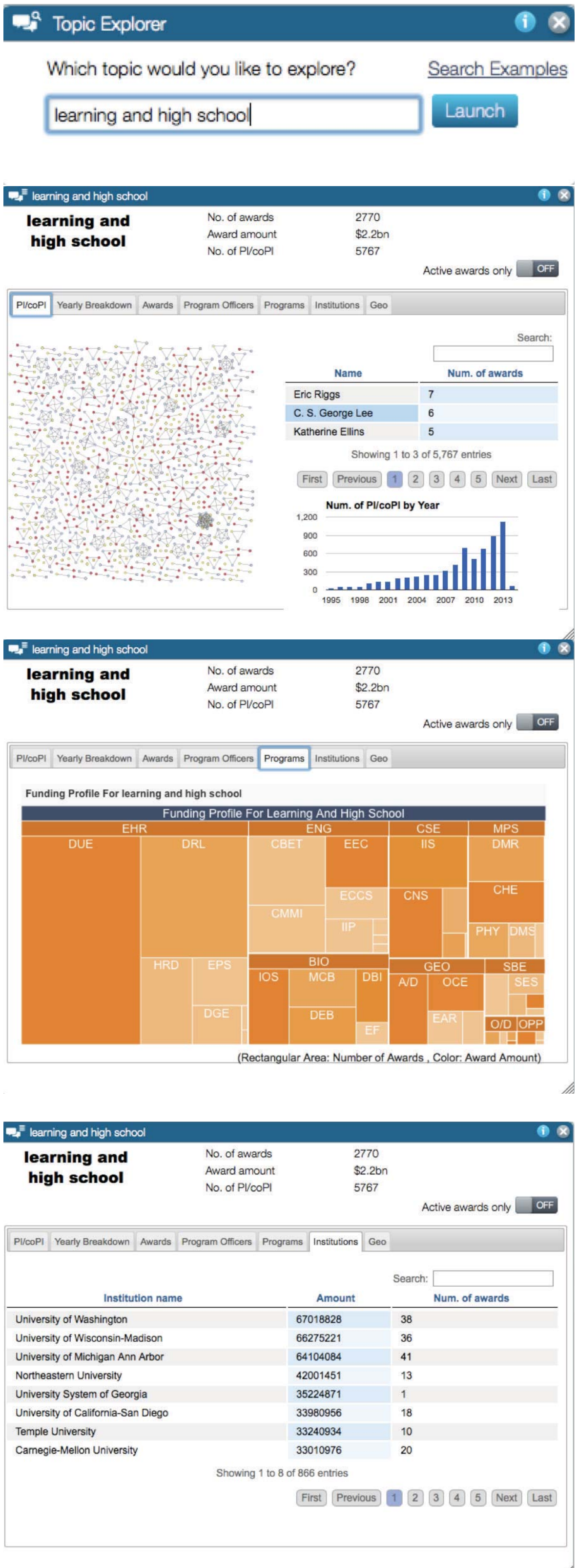

Fig. 4. Search service interacts with the rules engine and visualization services to provide end users with highly aggregated results as opposed to long list of results.

or self-explanatory to a layperson. This includes simple statistical graphics such as bar charts, line graphs, and pie charts as well as basic tables and lists. We also opted to include simple representations based on geographical maps, such as choropleth maps (Fig. 3), because they are part of the visual vocabulary of current print and digital media. However, some of our DIA2 data, such as the collaboration networks (CDVS), or multidimensional tables describing awards and programs, required more complex visual representations to fully visualize. For the former, we opted for node-link representations to show actors and their relations (Fig. 2), reasoning that this mimics conventions used in current social media platforms that many of our user personas are exposed to in their personal lives. This design choice turned out to be questionable—see Section 7—and needs further improvement.

Visualizing multidimensional datasets, such as all of the facets of a funding portfolio for a particular program, was a difficult challenge. Our current solution is to use interactive progressive refinement, where detail (and thus complexity) is gradually added to a visual representation as an effect of user interaction, such as filtering or pivoting on data. Figure 4 shows how a user starting from the Topic Explorer widget could progressively refine the search in order to get access to specific information and gain "deep insights." The search for the topic "learning and high school" opens a Topic Profile widget with several tabs. Each tab shows different information, and is interactive. For example, the first tab shows the collaboration network of PIs and Co-PIs working on this topic. The side table presents their names along with the number of NSF awards for each individual. The table is searchable, which enables users to retrieve specific individuals. Clicking an individual's name would bring up the People Profile widget with information about the person's collaborators, awards, program officers, institutional affiliation, and more. The Topic Profile widget also shows what NSF units fund awards in the area of "learning and high school." The treemap rectangles can be clicked for information about specific awards, PIs and POs within that specific organizational unit. The Topic Profile also shows a list of POs who manage awards in the area of "learning and high school," a list of award titles linked to abstracts, and a list of institutions that have received funding in this area. The Geo tab shows the geographical distribution of funding across U.S. states. The POs indicated this feature would be useful for the institution's accountability to local and federal government.

7 EVALUATION

The alpha version of DIA2 was tested with users from the NSF. The primary goal of the evaluation was to assess ease of learning. The following questions guided the evaluation: 1) Do users know how to use the interface – are they aware of what they can do, and how they can do it, without instructions? 2) Are users able to interpret the visualizations without explanations from the moderator? Our ideal research participant would be an NSF employee who had not previously seen a presentation and demo of the system. Without direct access to participants, we had to rely on gatekeepers at NSF to recruit participants for us. Five program officers (3 POs and 2 rotators) and two science assistants agreed to examine the interface and describe out loud their thoughts [27] as they tried to understand what the tool did and how to use it. The 3 POs all had more than 5 years' experience at NSF. Among the rotators and the science assistants, one each had less than 1 year experience at NSF and one between 1-2 years. Although the total sample size is sufficient for identifying usability issues that compromise ease of learning, each user group is not sufficiently represented so we are not able to draw comparative conclusions. With such small numbers of participants from each group, no clear patterns emerged during data analysis that differentiated among the groups. We used a moderated usability interview to collect data, following the active intervention method proposed by [11]. The research protocol started with the users viewing the start screen of DIA2 and talking out loud about their understanding of what they could do from here and how. After the initial screen, we allowed users to explore the system by starting the widgets they were interested in. When needed, we asked participants to perform some tasks, or to pursue questions that came up. Even though we asked participants to perform some tasks using DIA2, we chose not to collect quantitative metrics such as time on task [37] and instead to focus on users' cognitive processes for learning and understanding the interface. The session concluded with a demographics and System Usability Scale (SUS) survey [2, 37]. The moderated sessions were video recorded, and the recordings were analyzed in order to assess ease of learning and identify usability issues. We define a usability issue as any aspect of the interface that users did not readily understand or were unable to use. During data analysis, we focused on identifying statements and behaviors related specifically to ease or difficulty of learning the system and understanding the data visualizations. We coded these statements as positive or negative as recommended by Tullis and Albert [38] and organized them around each component of the system.

7.1 Evaluation Results

Overall, the system received positive feedback from users, one of whom made comments such as "this thing reads my mind" and "I feel it was designed for me." These are indicators that the users' mental models of their work with data were indeed reflected in the design. The average SUS score was 73.33 out of 100, which falls into the acceptable range [2]., Feedback on specific visualizations indicated areas of improvement.

Specifically, the **treemap visualization** is used repeatedly in DIA2 to show the allocation of funds and numbers of awards (proposals) across organizational units. None of our users seemed to be familiar with treemap visualizations prior to the testing. They used their knowledge of the organization's structure to infer the meanings associated with block size and color saturation on the treemap visualization. All users were able to figure it out after a few seconds of thinking. Even though a one-line legend was available at the bottom, users did not read it. Some users reported not noticing it, and others reported they enjoyed figuring things out by themselves and did not want to read the half-line legend. The users preferred the graphical parts of the interface and focused on them unless prompted to read the labels. This finding is interesting from two different points of view. First, our casual experts assume sufficient domain expertise and enjoy figuring things out by themselves, so they are unlikely to read any instructions, no matter how short. Second, the use of organizational units they were already familiar with seemed to facilitate their quick learning of the treemap visualization.

DIA2 also uses a number of traditional data representations such as **tables and bar charts**. Even though we identified a number of user requests related to sorting and pagination of information in tables, users, as expected, had no trouble understanding them. Users asked to sort information by additional columns, to be able to restrict the range of information to only awards that are currently active, and suggested scrolling instead of clicking through multiple pages of table rows.

Collaboration networks are another type of visualization frequently used in DIA2.. For example, a program officer can see not only a list of all the individuals and awards he or she has funded, but also a social network of all the individuals funded. We assumed that collaboration networks would be an interesting measure of a discipline's development and would be useful in helping program officers identify conflicts of interest. Close collaborators, for example, cannot serve on panels evaluating each other's proposals. However, such network representations were not readily understood by users. Labels for each node appear in the visualization only on mouse-over. The nodes are represented as small circles. Therefore, the networks failed to clearly and quickly communicate to users that the nodes represented people. Upon exploration, most users understood what the networks represented, although a couple of them needed some explanations from the moderator. Even so, the users were not sure what the links between nodes represented, or what the meaning of the color-coding was. They evaluated the visualization as "cool" and "interesting" but were not yet sure as to how they would use it in their daily work. This finding could be explained by the fact that NSF employees do not think of collaboration networks as part of their daily work. Instead, the organizational structure of programs, divisions and directorates is the predominant way to organize information. Also, the shape of the nodes failed to quickly communicate their nature. We learned that the network visualization needs to communicate affordance more clearly and include brief pointers and legends that can orient users.

Collectively, these findings show that the treemap visualization, which was completely new to users, made more sense to them than the relatively popular social network visualization. We explain this finding by reverting to this project's guiding concept, that of mental models. The way we used treemap visualizations in DIA2 was consistent with the users' mental model of how data is reported with respect to organization structure and therefore they could rely on that existing knowledge to make sense of the new information. However, because NSF staff members rarely thought in terms of collaborative research networks, seeing researchers represented this way made less sense to them than the treemap. Users showed interest in this new perspective and were open to its potential, but experienced difficulty learning it on their own.

The evaluation results suggest that even though users are able to learn how to use DIA2 without manuals or training, there are specific details that need to be improved in order for all visualizations to make sense at first exposure.

8 Lessons Learned

Sedlmair et al. [35] point to the specific characteristic of design studies as using visualization expertise to understand and build solutions that are able to address real-world problems faced by domain experts. DIA2 serves this purpose. The end users of our system require deep insights about their portfolio at a very high level of precision to be able to address various real-world policy concerns and direct funding. To this date, as far as we know, no other team in the world has managed to derive such deep insights into the real-world problems faced by users within a governmental agency such as the U.S. National Science Foundation. Our solution not only utilizes publicly available data, but DIA2 is being deployed directly inside the NSF firewalls. This level of impact comes with significant design challenges that we address next.

8.1 Designing with Limited Access

One of the most important and critical reasons why previous efforts to build a portfolio analysis system by external researchers (meaning not staff, employed by, or under contract of) the NSF is that research teams cannot have access to internal datasets directly. This made building a data-driven solution virtually impossible. One of the contributions of DIA2 is the realization that data-driven systems can be built as long as access to metadata schemas can be provided. The DIA2 team did not have direct access to the NSF data. In fact, we were never allowed to look at the data. Complicating this further, over the years, NSF has built several databases and systems, each containing slightly different fields for the same data. The schemas of these databases are not well documented or easily readable. The science assistants use a relatively complete Microsoft SQL database that connects to a financial system. However, due to confidentiality concerns, we were only given *metadata schema* access to two less complete databases—a SQL database and a series of XML data files. We spent a considerable amount of time understanding and bridging the two data sources. However, using our design process, as described in this paper, we were still able to understand the users, derive user requirements, build highly tailored solutions for the audience, and deploy this solution. Using a goal-directed design process, we are designing DIA2 to address a major national need in understanding the NSF's funding portfolio. This requires close collaboration and trust between the external researchers and users inside the NSF. Similarly, access to users themselves was limited and hard to obtain. We were fortunate to be able to create good working relationships with gatekeepers who mediated our access to a very busy population. Even though we were not able to conduct our formative research or evaluation using ideal methods, we were able to get information about our users' needs and goals that helped us design a system they find meaningful and useful.

DIA2 is in essence a great example of how to work within the legal framework of data at federal agencies while still delivering value using visual analytics. Among all our end users, science assistants have the most hands-on experiences with the real data, therefore, by working with the science assistants closely, we were able to "design in the dark" and reach the intended results without ever seeing the real data. A final critical component in our approach was adopting an agile development method of releasing early and often; our users on the other side of the wall (i.e. who had access to the confidential data) could then give us rapid informal feedback on the results to allow for changes.

8.2 Affordance is Innovation

The visualization and visual analytics community has at times a tendency to dismiss applied work as not innovative. The true value of visualizations or indeed visual analytics has to be in the new affordances its use offers to end-users. DIA2 takes on the challenge of providing insights at speed in a context where the stakes are high. From our user studies, we understand that in the design of systems like DIA2, it is extremely critical to select simple and useful representations of data rather than to strive for the creation of absolutely new algorithms and visualizations.

Furthermore, what is even more critical is to offer insights at high speed while reducing the cognitive burdens on the end-users. This enables them to perform many more analyses in more meaningful ways than before. Also, they are now able to ask more critical questions than were possible before. Our argument in this paper is that for user-focused systems like DIA2, ensuring that the end-users maximize on the value of the knowledge mining platform is more important than novelty in visual representations and analytics – *affordance is innovation*. This echoes findings by Sedlmair et al. [35] and is perhaps one of the most valuable implications for designing for casual experts. Using popular types of visual analytics, mapping them to the ways users think about their work, and employing progressive refinement are some of the techniques designers can employ when designing for "casual experts."

8.3 Fitting into Existing Organizational Ecosystems

One of the biggest challenges to introducing a system such as DIA2 into an environment like the U.S. NSF is that it needs to fit in with the organizational and cultural norms of that institution. Furthermore, even with the public data, the simple visuals and ability to mine massive amounts of data in an easy and intuitive way opens up the awards portfolio to a level of scrutiny that organizations need to prepare and plan for. It is true that such data are available publicly, but what is different is the ability to see the strengths and weaknesses of a program or organizational branch very simply. Furthermore, new systems such as DIA2 that are introduced into an organization must inevitably adapt to an existing ecology of both software—such as databases, management software, and search interfaces—as well as hardware—including server rooms, network architectures, and security systems. Adapting both the software and hardware aspects of DIA2 to the needs of end-users without losing on the ability to innovate scientifically is truly non-trivial.

9 Conclusion and Future Work

In this paper we have presented a design study of DIA2, a project designed to facilitate portfolio mining for the U.S. National Science Foundation program officers and assistants. In this system we have targeted a novel user population as well as a novel problem domain. Although a number of internal applications were available to users, none of them were designed with the user in mind; they were standardized packages modified for the users. As a consequence, the available solutions often proved inadequate, and adapting them to the users was hard for the designers as they did not understand the domain of the users. In our effort we had to start from scratch and our initial plan was to use novel and popular techniques currently in vogue and that had proved useful for a lot of other domains. As we learned more about how our users think about data, we revised our plans in favor of simpler but more useful techniques.. We wanted our system to provide new insights to the users but also support them in their tasks and reduce the time needed to respond to questions. The visualizations are a 'palette' of different kinds that are useful for understanding this domain and similar domains where organizational structure and function are largely in silos with only some integration across functions. Although our work does not contribute novel techniques or algorithms per se, the novelty of our work lies in the design approach, targeted domain, and the specific user group we name "casual experts." There are several design lessons learned from this design study such as how to design for specific organizational structures, and, how to translate mental models into design requirements and visualizations.

Our future work will focus on continuing to develop the DIA2 system in response to our end-users. We are also developing a community-facing version of the system that will help answer the same portfolio mining questions for our colleagues in the scientific community. We plan to continue evaluating the system and assessing its utility for each user group. Finally, we are highly interested in the concepts of casual experts and progressive refinement for visual analytics, and hope to continue exploring how to better accommodate these design constraints in future visual analytics and visualization systems.

Acknowledgment

This work was supported by the U.S. National Science Foundation awards TUES-1123108, TUES-1122609, and TUES-1123340. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] J. O. Bak, K. Nguyen, P. Risgaard, and J. Stage, "Obstacles to usability evaluation in practice: a survey of software development organizations," in *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*, 2008, pp. 23–32.

[2] A. Bangor, P. Kortum, and J.A. Miller, "Determining what individual SUS scores mean: Adding an adjective rating scale," *Journal of Usability Studies*, vol 4, no. 3, pp. 114-123, 2009.

[3] M. Bichier and K.-J. Lin, "Service-oriented computing," *Computer*, vol. 39, no. 3, pp. 99–101, 2006.

[4] R. Boyatzis, Transforming qualitative information: Thematic analysis and code development. Thousand Oaks, CA: Sage, 1998.

[5] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology,* vol. 3, no. 2, pp. 77-101, Jan. 2006.

[6] S. Chakrabarti, M. Van den Berg, and B. Dom, "Focused crawling: a new approach to topic-specific Web resource discovery," *Computer Networks*, vol. 31, no. 11, pp. 1623–1640, 1999.

[7] K. Charmaz, Constructing grounded theory. Sage, 2014.

[8] C. Chen, "Visualising semantic spaces and author co-citation networks in digital libraries," *Information processing & management*, vol. 35, no. 3, pp. 401–420, 1999.

[9] A. Cooper, R. Reimann, and D. Cronin, *About Face 3: The Essentials of Interaction Design*, 3rd ed. Wiley, 2007.

[10] M. Diligenti, F. Coetzee, S. Lawrence, C. L. Giles, and M. Gori, "Focused Crawling Using Context Graphs.," in *VLDB*, 2000, pp. 527–534.

[11] J. S. Dumas and J. C. Redish, *A Practical Guide to Usability Testing*. Intellect Books, 1999.

[12] T. Erl, Service-Oriented Architecture (SOA): Concepts, Technology and Design. Upper Saddle River, NJ: Prentice Hall PTR. 2005.

[13] S. Ghani, N. Elmqvist, and D. S. Ebert, "MultiNode-Explorer: A Visual Analytics Framework for Generating Web-based Multimodal Graph Visualizations," *Proc. of EuroVA'12*, pp. 67–71, 2012.

[14] S. Ghani, B. C. Kwon, S. Lee, J. S. Yi, and N. Elmqvist, "Visual analytics for multimodal social network analysis: A design study with social scientists," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, no. 12, pp. 2032–2041, 2013.

[15] T. M. Green, W. Ribarsky, and B. Fisher, "Building and applying a human cognition model for visual analytics," *Information visualization*, vol. 8, no. 1, pp. 1–13, 2009.

[16] T. M. Green, W. Ribarsky, and B. Fisher, "Visual analytics for complex concepts using a human cognition model," in *Visual Analytics Science and Technology, 2008. VAST'08. IEEE Symposium on*, 2008, pp. 91–98.

[17] V. Grigoreanu and M. Mohanna, "Informal cognitive walkthroughs (ICW): paring down and pairing up for an agile world," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2013, pp. 3093–3096.

[18] J. Grudin, "Obstacles to participatory design in large product development organizations," *Participatory design: principles and practices*, pp. 99–119, 1993.

[19] K. Holtzblatt and J. Sandra, "Contextual inquiry: a participatory technique for system design," *Participatory design: principles and practice*. Hillsdale, NJ: Erlbaum, 1993.

[20] W. W. Hood and C. S. Wilson, "The literature of bibliometrics, scientometrics, and informetrics," *Scientometrics*, vol. 52, no. 2, pp. 291–314, 2001.

[21] J. D. Mackinlay, R. Rao, and S. K. Card, "An organic user interface for searching citation links," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1995, pp. 67–73.

[22] K. Madhavan, M. Vorvoreanu, N. Elmqvist, A. Johri, N. Ramakrishnan, G. A. Wang, and A. F. McKenna, "Portfolio Mining," *IEEE Computer*, vol. 45, no. 10, pp. 95–99, 2012.

[23] P. D. Manuel and J. AlGhamdi, "A data-centric design for n-tier architecture," *Information Sciences*, vol. 150, no. 3, pp. 195–206, 2003.

[24] D. Modjeska, V. Tzerpos, P. Faloutsos, and M. Faloutsos, "BIVTECI: A bibliographic visualization tool," in *Proceedings of the 1996 conference of the Centre for Advanced Studies on Collaborative research*, 1996, p. 28.

[25] J. Nielsen, "10 Heuristics for User Interface Design." [Online]. Available: http://www.useit.com/papers/heuristic/heuristic_list.html.

[26] D. Norman, *The design of everyday things*. Basic books, 2002.

[27] C. Lewis, Using the" thinking-aloud" method in cognitive interface design. IBM TJ Watson Research Center, 1982.

[28] Z. Liu and J. T. Stasko, "Mental models, visual reasoning and interaction in information visualization: A top-down perspective," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 16, no. 6, pp. 999–1008, 2010.

[29] Q. Liu, M. Vorvoreanu, K. P. Madhavan, and A. F. McKenna, "Designing discovery experience for big data interaction: a case of web-based knowledge mining and interactive visualization platform," in *Design, User Experience, and Usability. Web, Mobile, and Product Design*, Springer, 2013, pp. 543–552.

[30] M. Ogawa and K.-L. Ma, "code_swarm: A design study in organic software visualization," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 15, no. 6, pp. 1097–1104, 2009.

[31] S. E. Poltrock and J. Grudin, "Organizational obstacles to interface design and development: two participant-observer studies," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 1, no. 1, pp. 52–80, 1994.

[32] Z. Pousman, J. T. Stasko, and M. Mateas, "Casual information visualization: Depictions of data in everyday life," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 13, no. 6, pp. 1145–1152, 2007.

[33] M. Sedlmair, A. Frank, T. Munzner, and A. Butz, "RelEx: Visualization for actively changing overlay network specifications," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, no. 12, pp. 2729–2738, 2012.

[34] M. Sedlmair, P. Isenberg, D. Baur, and A. Butz, "Information visualization evaluation in large companies: Challenges, experiences and recommendations," *Information Visualization*, vol. 10, no. 3, pp. 248–266, 2011.

[35] M. Sedlmair, M. Meyer, and T. Munzner, "Design study methodology: Reflections from the trenches and the stacks," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, no. 12, pp. 2431–2440, 2012.

[36] B. Shneiderman and M. Wattenberg, "Ordered treemap layouts," in *Information Visualization, IEEE Symposium on*, 2001, pp. 73–73.

[37] M. Tory and T. Moller, "Human factors in visualization research," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 10, no. 1, pp. 72–84, 2004.

[38] T. Tullis and W. Albert, Measuring the user experience: Collecting, analyzing, and presenting usability metrics. Burlington, MA: Morgan Kaufman, 2008

[39] I. H. Witten, G. W. Paynter, E. Frank, C. Gutwin, and C. G. Nevill-Manning, "KEA: Practical automatic keyphrase extraction," in *Proceedings of the fourth ACM conference on Digital libraries*, 1999, pp. 254–255.

[40] J. Wood, D. Badawood, J. Dykes, and A. Slingsby, "BallotMaps: Detecting name bias in alphabetically ordered ballot papers," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 12, pp. 2384–2391, 2011.